Hans J. Stetter

Richardson-extrapolation and optimal estimation

Aplikace matematiky, Vol. 13 (1968), No. 2, 187–190

RICHARDSON-EXTRAPOLATION AND OPTIMAL ESTIMATION

H. J. STETTER

Although the principle of Richardson-extrapolation is 40 years old and although it has been applied with tremendous success in many diverse applications, with some numerical analysts there have remained uneasy feelings with respect to its mathematical foundations: It is true that the existence of an asymptotic expansion in powers of the discretization parameter can be proved for a wide class of problems (see e.g. [1]) but only in the case of the trapezoidal discretization of definite integrals the coefficients of the expansion are practically accessible so that an error estimate can be derived (see e.g. [2], [3]). Under certain conditions (which are always satisfied if the discretization is "fine" enough but which cannot be explicitly verified) the monotony of the Richardson-extrapolates permits the derivation of upper and lower bounds for the approximated values (see e.g. [4]). This seems unsatisfactory to them.

If one regards other methods in numerical analysis, however, one finds that most of these rest on no firmer ground. A bound on some high derivative of a function determined from numerical data e.g., is as inaccessible in most cases as are the coefficients of an asymptotic expansion for a discretization error. Therefore, increasing emphasis has been placed lately on lines of thought which fall under the pattern "estimation" rather than "approximation". (These ideas are stressed and treated in many papers by BABUŠKA, see [5] for a very general juxtaposition of the two lines of thought.)

In the following we will try to show that — under reasonable conditions — Richardson-extrapolation is an optimal estimation method. This should serve as a further motivation for its extended use.

In order to keep the proof short and transparent we will only consider the very simplest case: Polynomial extrapolation from two computed values. The more general cases can be treated similarly and yield analogous results.

Consider a problem with solution $\hat{y}$ (a real number for simplicity) and a particular discretization with a solution family $\eta(h)$, $h > 0$ is the discretization parameter. The admissible values of h constitute the set H. $\eta(h)$ is assumed to be computable without round-off, i.e. the round-off is considered negligible within the present context.

Let the existence of the following asymptotic law be known

$$(1) \qquad \eta(h) = \hat{y} + eh^p + r(h), \quad \text{with} \quad r(h) = O(h^{p+1}) \quad \text{for} \quad h \to 0,$$

where $p > 0$ is a known number but the value of e is inaccessible (e.g. a complicated expression in derivatives of the function of which $\hat{y}$ is a value at a given point). Relation (1) means the following: There are two well-defined (but unknown) numbers $\underline{r}$ and $\bar{r}$, $\underline{r} \leqq \bar{r}$, such that

$$\hat{y} \in Y(h) := \eta(h) - eh^p - [\underline{r}, \bar{r}]\, h^{p+1} \tag{2}$$

holds for all $h \in H$. (The expression on the right-hand side is an interval, according to the notation of interval analysis, e.g. [6] or [7].)

Assume that η has been computed for h_1 and h_2, $0 < h_1 < h_2$. We want to obtain from $\eta(h_1)$ and $\eta(h_2)$ the best possible estimate for $\hat{y}$ under the given information. For this purpose we consider the one-parameter set of linear combinations

$$y(c) := (1 + c)\, \eta(h_2) - c\, \eta(h_1) \tag{3}$$

and choose c_0 such that

$$\max_{\hat{y} \in Y(h_1) \cap Y(h_2)} |y(c_0) - \hat{y}| = \min_c \max_{\hat{y} \in Y(h_1) \cap Y(h_2)} |y(c) - \hat{y}|\,. \tag{4}$$

Then the value of $y(c_0)$ is the optimal estimate for $\hat{y}$ under the given information according to the general theory of estimation (see e.g. [5]).

Of course, c_0 will depend on the unknown values of e, $\underline{r}$ and $\bar{r}$. However, we will now show that these values need only satisfy a weak relation in oider that c_0 become independent of the parameters and equal to the value obtained from (polynomial) Richardson-extrapolation.

Theorem. *If*

$$|e|\,(h_1^p - h_2^p) \geqq \left|\frac{\bar{r} + \underline{r}}{2}\right| (h_1^{p+1} - h_2^{p+1}) + \frac{\bar{r} - \underline{r}}{2} (h_1^{p+1} + h_2^{p+1}) \tag{5}$$

then

$$c_0 = \frac{h_2^p}{h_1^p - h_2^p}\,. \tag{6}$$

Proof. From (1) and (2) we have

$$\begin{aligned}
\hat{y} &= (1 + c)\,(\eta(h_2) - eh_2^p - r(h_2)) - c(\eta(h_1) - eh_1^p - r(h_1)) \\
&= y(c) - e((1 + c)\, h_2^p - ch_1^p) - (1 + c)\, r(h_2) + cr(h_1) \\
&\in y(c) + e(ch_1^p - (1 + c)\, h_2^p) \\
&\quad + [c\underline{r}h_1^{p+1} - (1 + c)\, \bar{r}h_2^{p+1}, c\bar{r}h_1^{p+1} - (1 + c)\, \underline{r}h_2^{p+1}] \\
&= y(c) + e(ch_1^p - (1 + c)\, h_2^p) + \frac{\bar{r} + \underline{r}}{2} (ch_1^{p+1} - (1 + c)\, h_2^{p+1}) \\
&\quad + \left[-\frac{\bar{r} - \underline{r}}{2} (ch_1^{p+1} + (1 + c)\, h_2^{p+1}),\ \frac{\bar{r} - \underline{r}}{2} (ch_1^{p+1} + (1 + c)\, h_2^{p+1})\right].
\end{aligned}$$

This implies

$$(7) \qquad |\hat{y} - y(c)| \leqq |e|\,|ch_1^p - (1 + c)\,h_2^p| + \left|\frac{\bar{r} + \underline{r}}{2}\right| |ch_1^{p+1} - (1 + c)\,h_2^{p+1}|$$

$$+ \frac{\bar{r} - \underline{r}}{2} |ch_1^{p+1} + (1 + c)\,h_2^{p+1}| = \max_{\hat{y} \in Y(h_1) \cap Y(h_2)} |\hat{y} - y(c)|$$

since the equality may actually occur under our information.

The three terms on right-hand side of (7) vanish at

$$c_0 := \frac{h_2^p}{h_1^p - h_2^p}, \quad c_0' := \frac{h_2^{p+1}}{h_1^{p+1} - h_2^{p+1}}, \quad c_0'' := -\frac{h_2^{p+1}}{h_1^{p+1} + h_2^{p+1}},$$

resp. where $c_0'' < c_0' < c_0$ due to $h_2 < h_1$ and $p > 0$.

It is evident that a unique minimum will occur at c_0 if the derivative with respect to c of the first term is larger than the sum of the derivatives of the other two (for equality we have a minimal interval $[c_0', c_0]$). But this is equivalent to the assertion of the theorem as is seen from (7).

Remarks. 1. Condition (5) is only unessentially stronger than

$$|e|\,h_i^p \geqq \max\left(|\bar{r}|, |\underline{r}|\right) h_i^{p+1}, \quad i = 1, 2.$$

Thus it requires that for the values of h used in the computation the error term in the asymptotic law should be smaller in absolute value than the expansion term in order that Richardson-extrapolation provide an optimal estimate. This is a very natural requirement, in any reasonable application it will be automatically satisfied for values of h_i which would be practically used for computation, except in cases where $e = 0$ or is very small by coincidence.

2. It should be quite clear that any additional information, e.g. on the relative signs of e and $r(h)$, may change the situation as the equality in (7) need no longer be attainable. However, such information can usually be obtained only from computing more than two values $\eta(h_i)$ (or from deeper analysis of the problem which we have ruled out).

3. In the case of a longer asymptotic expansion (1) the optimality of the corresponding multiple polynomial Richardson-extrapolation from several values $\eta(h_i)$ is obtained under similar conditions. The details become more cumbersome.

4. Instead of regarding only linear combinations (3) of the computed values we could have considered expressions in $\eta(h_1)$ and $\eta(h_2)$ which correspond to rational extrapolation (see [3]). In the case of only two computed values, at least, the optimal estimate is again given by the expression for Richardson-extrapolation. The investigations become unmanageable for more than two $\eta(h_i)$.

It has been the purpose of this paper to show that Richardson-extrapolation reveals itself as an optimal estimation principle if the line of thought suggested in [5] is applied. This should be considered as further evidence of the basic importance of the method.

References

[1] *H. J. Stetter:* Asymptotic expansions for the error of discretizations algorithms for non-linear functional equations, Numer. Math. *7* (1965), 18—31.

[2] *R. Bulirsch:* Bemerkungen zur Romberg-Integration, Numer. Math. *6* (1964), 6—16.

[3] *R. Bulirsch* and *J. Stoer:* Über Fehlerabschätzung und Extrapolation mit rationalen Funktionen bei Verfahren vom Richardson-Typus, Numer. Math. *6* (1964), 413—427.

[4] *R. Bulirsch* and *J. Stoer:* Asymptotic upper and lower bounds for results of extrapolation methods, Numer. Math. *8* (1966), 93—104.

[5] *J. R. Rice:* Numerical Analysis — approximation or estimation?, SICNUM news letter *2* (1967), No. 2.

[6] *R. E. Moore:* Interval analysis, Prentice-Hall, 1966.

[7] *N. Apostolatos* and *U. Kulisch:* Grundlagen einer Maschinenintervallarithmetik, Computing, *2* (1967), 89—104.

H. J. Stetter, Institut für Numerische Mathematik, TH Wien, A—1040, Wien, Österreich.